



# Modeling the Active and Idle Durations of Network Hosts

Soumyo Moitra  
smoitra@cert.org  
FloCon 2015



Report Documentation Page				Form Approved OMB No. 0704-0188	
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE <b>JAN 2015</b>		2. REPORT TYPE		3. DATES COVERED <b>00-00-2015 to 00-00-2015</b>	
4. TITLE AND SUBTITLE <b>Modeling the Active and Idle Durations of Network Hosts</b>				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) <b>Carnegie Mellon University, Software Engineering Institute, Pittsburgh, PA, 15213</b>				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT <b>Approved for public release; distribution unlimited</b>					
13. SUPPLEMENTARY NOTES <b>FloCon 2015, Portland, OR, January 12-15, 2015.</b>					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT <b>Same as Report (SAR)</b>	18. NUMBER OF PAGES <b>11</b>	19a. NAME OF RESPONSIBLE PERSON
a REPORT <b>unclassified</b>	b ABSTRACT <b>unclassified</b>	c THIS PAGE <b>unclassified</b>			

---

This material is based upon work funded and supported by SEI Line Funding under Contract No. FA8721-05-C-0003 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center sponsored by the United States Department of Defense

NO WARRANTY. THIS CARNEGIE MELLON UNIVERSITY AND SOFTWARE ENGINEERING INSTITUTE MATERIAL IS FURNISHED ON AN "AS-IS" BASIS. CARNEGIE MELLON UNIVERSITY MAKES NO WARRANTIES OF ANY KIND, EITHER EXPRESSED OR IMPLIED, AS TO ANY MATTER INCLUDING, BUT NOT LIMITED TO, WARRANTY OF FITNESS FOR PURPOSE OR MERCHANTABILITY, EXCLUSIVITY, OR RESULTS OBTAINED FROM USE OF THE MATERIAL. CARNEGIE MELLON UNIVERSITY DOES NOT MAKE ANY WARRANTY OF ANY KIND WITH RESPECT TO FREEDOM FROM PATENT, TRADEMARK, OR COPYRIGHT INFRINGEMENT.

This material has been approved for public release and unlimited distribution.

The Government of the United States has a royalty-free government-purpose license to use, duplicate, or disclose the work, in whole or in part and in any manner, and to have or permit others to do so, for government purposes pursuant to the copyright license under the clause at 252.227-7013 and 252.227-7013 Alternate I.

This material was prepared for the exclusive use of FloCon 2015 attendees and may not be used for any other purpose without the written consent of [permission@sei.cmu.edu](mailto:permission@sei.cmu.edu).

DM-0001657

# Introduction

---

Important to understand network behavior of hosts

Durations active and idle by host type

Patterns important for Situational Awareness

Baselining to detect anomalies

Decide whether a host should be in the inventory

# Objectives of the Analysis

---

Distributions of the durations of active and idle times

Insights into different behaviors

Two metrics:

Probability of a host being active after a period of idleness

Conditional probability of a host becoming active within a time horizon  
Given it has been idle for some time

# Methodology

---

Flow data from the public domain

[\(http://tools.netsa.cert.org/silk/referencedata.html\)](http://tools.netsa.cert.org/silk/referencedata.html)

SiLK (CERT/SEI) and Unix Tools

Spreadsheets

Focus on web servers initially

Methodology applicable to all types of hosts

# References

---

- Bhattacharya, R. N. and Waymire, E. C. (2009) Stochastic Processes with Applications. SIAM.
- Brostrom, G. (2012) Event History Analysis with R. CRC Press.
- Crovella, M. and Krishnamurthy, B. (2006) Internet Measurement. John Wiley & Sons.
- Hayden, L. (2010) IT Security Metrics. McGraw Hill.
- Lawless, J. F. (2002) Statistical Models and methods for Lifetime Data. Wiley.
- Maindonald, J. and Braun, W. J. (2010) Data Analysis and Graphics using R: An Example-Based Approach. Cambridge University Press.
- Mills, M. (2011) Introducing Survival and Event History Analysis. Sage.
- Rausland, M. and Heyland, A. (2003) System Reliability Theory. Wiley.
- Ross, S. (2014) Applied Probability Models. Academic Press.
- Snyder, D. L. and Miller, M. I. (2011) Random Point Processes in Time and Space. Springer.

# Analysis

---

Time series of network flows – out traffic

Time window = 23 hours

Time scale (bin size) = 1 hour

Convert volumes to a 0/1 series (1 => active)

Compute the durations of active and idle times

Plot the frequency distributions



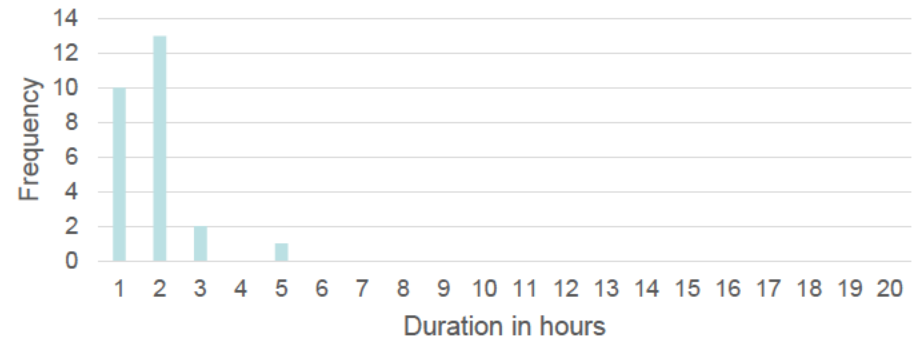
# Durations from Flows (Hypothetical)

---

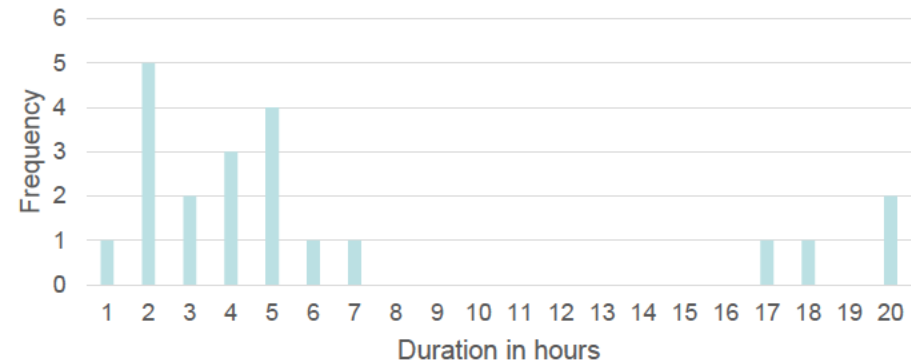
Flows from rwcount	Conversion to 1/0	<u>I</u>	<u>U</u>
123	1		
456	1		
789	1	3	
0	0		
0	0		2
234	1		
90	1	2	
0	0		
0	0		
0	0		
0	0		4
55	1	1	
0	0		1
99	1		

# Results

Distribution of active durations



Distribution of idle durations



# Discussion

---

## Active durations

- Very compact (low variation – narrower than Poisson)

- Mean = 1.8

- Weibull?

## Idle durations

- Long tail or two populations

- Issues with estimating the metrics

- Censoring/Truncation problems

## Future Work

- Need much longer time series

- Need to estimate the metrics with more data sets

- Correct for biases

- Compare across different host types

- Effects of varying the time scales, time windows and time horizons



**Thank you!**

**Questions/comments?**

